Causal inference in the relative survival framework

Elisavet Syriopoulou [1],
Mark J. Rutherford[1], Paul C. Lambert[1,2]

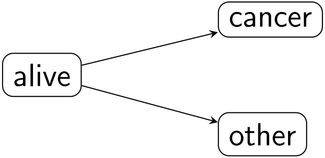
[1]Biostatistics Research Group, Department of Health Sciences, University of Leicester, Leicester, UK
[2]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

es303@le.ac.uk

12 July 2018

Background

In cancer epidemiology, competing risks are common.



We can estimate one of the following:

- Cause-specific mortality
- Excess mortality

Problems may arise by the inaccuracy or non-availability of the cause of death information.

We focus on estimating excess mortality that does not require the information on cause of death.

Excess mortality and Relative survival

Excess mortality

$$\text{excess mortality} = \text{all-cause mortality} - \text{expected mortality}$$

$$\lambda(t) = h(t) - h^*(t)$$

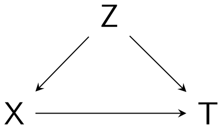
Relative survival

$$\text{relative survival ratio} = \frac{\text{all-cause survival proportion}}{\text{expected survival proportion}}$$

$$R(t) = \frac{S(t)}{S^*(t)} \qquad S(t) = S^*(t)R(t)$$

The expected survival is considered to be known and is obtained from available life tables on a comparable population.

Average causal effects



- Let T_i^0 and T_i^1 be the outcomes (time-to-event) that would be observed if patient i was unexposed ($X = 0$) or exposed ($X = 1$) respectively (potential outcomes).
- For each patient, one of the potential outcomes is counterfactual and will never be realised.
- Estimation therefore focuses on estimating averages, such as the average causal difference

$$E(T^1) - E(T^0)$$

$$Pr(T^1 < t) - Pr(T^0 < t)$$

Standardised survival function

The standardised survival function at X = x, is defined as

$$E\left[S(t|X=x,Z)\right] = \frac{1}{N}\sum_{i=1}^{N}\widehat{S}(t|X=x,z_i).$$

- Fit a survival model, such as Cox model or flexible parametric model.
- Obtain survival predictions for each individual in the population.
- Calculate an average of the survival predictions.

Difference of standardised survival functions

If interested in **all-cause** survival:

$$E\left[S(t|X=1,Z)\right]-E\left[S(t|X=0,Z)\right]$$
$$E\left[S^{*}(t|X=1,Z)R(t|X=1,Z)\right]-E\left[S^{*}(t|X=0,Z)R(t|X=0,Z)\right]$$

If interested in **cancer-related** survival:

$$E\left[R(t|X=1,Z)\right]-E\left[R(t|X=0,Z)\right]$$

- This difference will refer to a hypothetical world where the cancer of interest is the only possible cause of death.
- However, it can be a very useful measure when comparing populations!
- Average causal difference under assumptions: conditional exchangeability, consistency, positivity, no interference, well-defined exposures.

Standardising to a subset of the study population

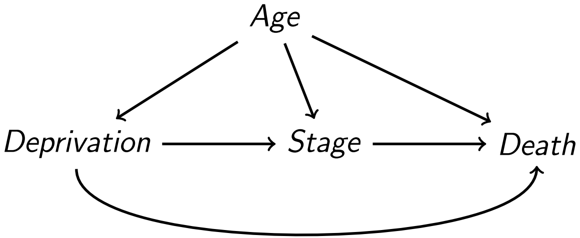
By standardising to the covariate distribution of the exposed:

$$E\left[R(t|X=1, Z_1)\right] = \frac{1}{N_1}\sum_{i=1}^{N_1} \widehat{R}(t|X=1, z_{1i})$$

$$E\left[R(t|X=0, Z_1)\right] = \frac{1}{N_1}\sum_{i=1}^{N_1} \widehat{R}(t|X=0, z_{1i}),$$

where N_1 is the proportion of patients within the exposed.

Mediation analysis



How much of the differences between deprivation groups can be explained by differences at the stage distribution?

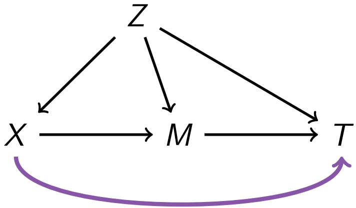
How many deaths could be avoided if the most deprived patients had the same relative survival as the least deprived? or same stage distribution?

Natural Direct Effects

For each level of Z:

$$NDE = E\left[R(t|X=1, M_1)\right] - E\left[R(t|X=0, M_1)\right]$$

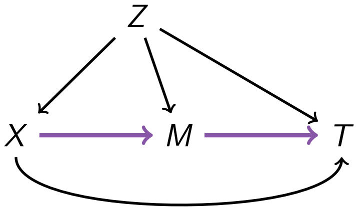
Note here the exposed have their own mediator distribution, but the unexposed have the mediator distribution of the exposed.



Natural Indirect Effects
For each level of Z:

$$NIE = E\left[R(t|X=1, M_1)\right] - E\left[R(t|X=1, M_0)\right]$$

Gives the effect of the mediator in the exposed.



Real world impact - Avoidable deaths

- The predicted number of deaths within strata $Z = z$ for exposure $X = 1$:

$$D_1(t|X=1, M_1) = N_{Z=z,X=1} \times \left(1 - E\left[S^*(t|X=1)R(t|X=1, M_1)\right]\right)$$

Real world impact - Avoidable deaths

- The predicted number of deaths within strata $Z = z$ for exposure $X = 1$:

$$D_1(t|X=1, M_1) = N_{Z=z,X=1} \times (1 - E\left[S^*(t|X=1)R(t|X=1, M_1)\right])$$

- The expected number of deaths if the exposed group had the same survival as the unexposed is

$$D_{1|0}(t|X=1, M_1) = N_{Z=z,X=1} \times (1 - E\left[S^*(t|X=0)R(t|X=0, M_1)\right])$$

Real world impact - Avoidable deaths

- The predicted number of deaths within strata $Z = z$ for exposure $X = 1$:

$$D_1(t|X = 1, M_1) = N_{Z=z,X=1} \times (1 - E\left[S^*(t|X = 1)R(t|X = 1, M_1)\right])$$

- The expected number of deaths if the exposed group had the same **(relative)** survival as the unexposed is

$$D_{1|0}(t|X = 1, M_1) = N_{Z=z,X=1} \times (1 - E\left[S^*(t|X = 0)R(t|X = 0, M_1)\right])$$

$$D^c_{1|0}(t|X = 1, M_1) = N_{Z=z,X=1} \times (1 - E\left[S^*(t|X = 1)\mathbf{R(t|X = 0, M_1)}\right])$$

Real world impact - Avoidable deaths

- The predicted number of deaths within strata $Z = z$ for exposure $X = 1$:

$$D_1(t|X=1,M_1) = N_{Z=z,X=1} \times (1 - E\left[S^*(t|X=1)R(t|X=1,M_1)\right])$$

- The expected number of deaths if the exposed group had the same **(relative)** survival as the unexposed is

$$D_{1|0}(t|X=1,M_1) = N_{Z=z,X=1} \times (1 - E\left[S^*(t|X=0)R(t|X=0,M_1)\right])$$

$$D^c_{1|0}(t|X=1,M_1) = N_{Z=z,X=1} \times (1 - E\left[S^*(t|X=1)\boldsymbol{R(t|X=0,M_1)}\right])$$

- The avoidable deaths are therefore,

$$AD(t|X=1,M_1) = D_1(t|X=1,M_1) - D^c_{1|0}(t|X=1,M_1)$$

Real world impact - Avoidable deaths

- The predicted number of deaths within strata $Z = z$ for exposure $X = 1$:

$$D_1(t|X=1, M_1) = N_{Z=z,X=1} \times (1 - E\left[S^*(t|X=1)R(t|X=1, M_1)\right])$$

- The expected number of deaths if the exposed group had the same **(relative)** survival as the unexposed is

$$D_{1|0}(t|X=1, M_1) = N_{Z=z,X=1} \times (1 - E\left[S^*(t|X=0)R(t|X=0, M_1)\right])$$

$$D^c_{1|0}(t|X=1, M_1) = N_{Z=z,X=1} \times (1 - E\left[S^*(t|X=1)\boldsymbol{R(t|X=0, M_1)}\right])$$

- The avoidable deaths are therefore,

$$AD(t|X=1, M_1) = D_1(t|X=1, M_1) - D^c_{1|0}(t|X=1, M_1)$$

- To get the total avoidable deaths we sum over all strata, Z.

$$AD(t, X=1, M_1) = \sum_{Z_1} AD(t|X=1, M_1)$$

Application

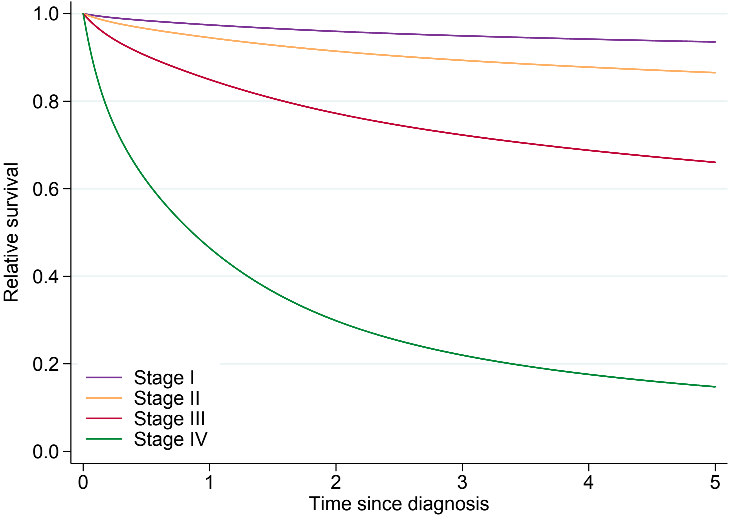
- We analysed data on colon cancer in England.
- Fitted a flexible parametric survival model that uses restricted cubic splines to model the baseline excess hazard:
    - Age at diagnosis as a continuous and non-linear variable
    - Time-dependent effects for deprivation and age
    - Interaction between deprivation and age
- Performed a period analysis with a period window from 2012 to 2013 to ensure that the estimates are accurate for those recently diagnosed.
- Multiple imputations methods for missing data at stage at diagnosis.
- Background mortality was incorporated in the model using life tables.

Relative survival and stage by deprivation (females)

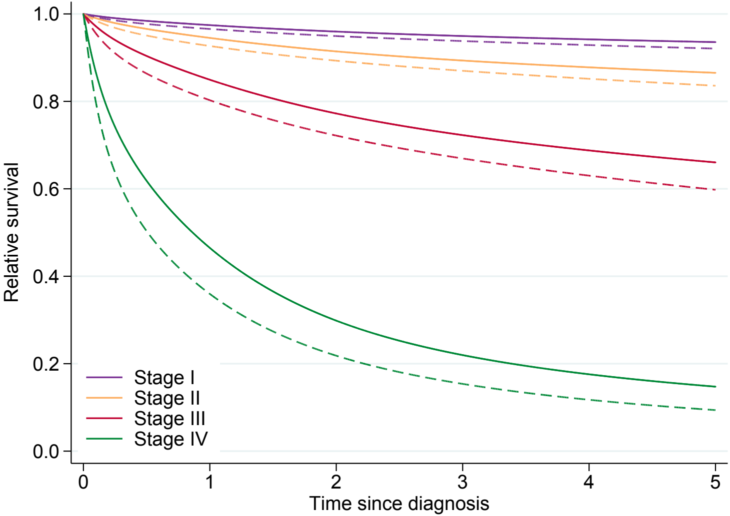
Age at Diagnosis	5-year Relative Survival	
	Least deprived	*Most deprived*
55	68.2	59.6
65	**66.5**	**56.6**
75	62.8	54.3
85	46.0	41.0

Stage at Diagnosis	Proportion of patients	
	Least deprived	*Most deprived*
I	14.0	13.4
II	29.9	30.5
III	28.2	25.9
IV	**27.9**	**30.2**

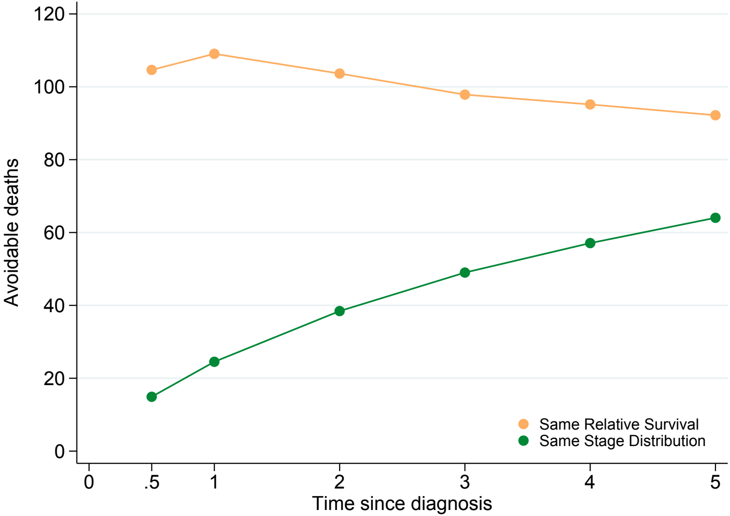
Relative survival by stage - least deprived



Relative survival by stage - least and most deprived



Avoidable deaths



Conclusions

- Relative survival controls for different background mortality across populations and enables fair comparisons.
- Avoidable deaths help quantify the impact of eliminating cancer inequalities in the real world.
- By applying a causal approach in cancer epidemiology we can make conclusions on the causal structure of variables and understand the underlying determinants of inequalities.
- Additional assumptions need to hold for the mediation analysis framework.
- Future work:
  - Model the mediator instead of stratification.
  - Extend software to estimate avoidable deaths at more than one time point.

Selected References

Pohar Perme, M., Stare, J. & Estève, J. On Estimation in Relative Survival. *Biometrics, Blackwell Publishing Inc*, 68, 113-120, 2012.

Royston, P. and Lambert, P.C. *Flexible parametric survival analysis in Stata: Beyond the Cox model*. Stata Press, 2011.

Sjölander A. Regression standardization with the R package stdReg. *European Journal of Epidemiology,* 31(6):563-574, 2016.

VanderWeele T.J. Causal mediation analysis with survival data. *Epidemiology,* 22(4):582-585, 2011.

Brenner, H., Gefeller, O. and Hakulinen, T. Period analysis for 'up-to-date' cancer survival data: theory, empirical evaluation, computational realisation and applications. *European Journal of Cancer,* 40, 326-335 ,2004.

Spika D., Rachet B., Bannon F., Woods L., Maringe C., Bonaventure A., et al. Life tables for the concord-2 study. Available from: http://csg.lshtm.ac.uk/life-tables.